Faculdade de Medicina da Universidade do Porto
Faculdade de Ciências da Universidade do Porto

2º Ciclo de Estudos
Em Informática Médica

Evidence of Temporal Bayesian Networks applications for health-related problems: a systematic review

Tiago Fernando da Silva Martins

Dissertation submitted for the degree of Master in Medical Informatics

Supervisors: Pedro Pereira Rodrigues

Cláudia Dias

Porto, September 2017

Rua Doutor Plácido da Costa | 4200-450 Porto . Portugal
Tel: +351 225 513 622 | Fax: +351 225 513 623
http://mim.med.up.pt . mim@med.up.pt

Acknowledgements

I want to begin this section with a quote that stands out in all my achievements in life: "When they say you can't, then you have to".

Foremost, I would like to express my sincere gratitude to my advisor Prof. Pedro Pereira Rodrigues for the continuous support on my master's thesis, for his immense knowledge and for the encouragement whenever I got unmotivated.

Besides my advisor, I would like to thank co-advisor Cláudia Dias for the continuous willingness to help, always providing wise advices and large doses of patience.

Finally, I must express my very profound gratitude to my parents and brother, for providing me with unfailing support, continuous encouragement and, above all, for all the unconditional love throughout my entire life, including the process of writing this thesis. I wouldn't be able to do it without you.

This study was performed under the scope of Project "NORTE-01-0145-FEDER-000016" (NanoSTIMA) is financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

Abstract

Aim: The aim of our study was to systematically review the literature on temporal Bayesian networks and its relation with clinical outcomes.

Background: Medical knowledge is growing at a fiery rate. While the obtainability of pertinent data has the potential to make the task of diagnosis more accurate, it is also progressively overwhelming for physicians to adapt. The wide range of studies addressing TBN models and applications available in the literature legitimates a formal systematic review of TBN applications to medical problems.

Methods: This systematic review was conducted in accordance with PRISMA guidelines. Search was conducted on multiple online databases, including PubMed, ISI Web of Knowledge, SCOPUS and DBLP, with no date or study design restrictions. We applied standard systematic review methodology for the identification of eligible studies, data extraction and assessment of risk of biases.

Results: From 8761 studies, 42 were deemed eligible for inclusion, and evidence on the benefits TBN's use was obtained. Twenty three (55%) studies addressed DBN as applied technique, sixteen (38%) approached the cardiovascular system and 28 (67%) studies were derived for prognosis. The analysis showed that the majority of studies found benefits on the use of TBN. Significant heterogeneity between was found between studies outcomes making it impossible to perform meta-analysis.

Discussion: This study had the ambition to perform a systematic review, arising novel data and content regarding temporal Bayesian network's application on health-related problems. There are a considerable number of published studies that investigate the benefits of temporal Bayesian networks applied on health-related problems. These results may encourage the use of TBN, given the shown enhanced performance. The resulting evidence and/or models might be used in a future development of clinical decision support systems.

Keywords: temporal Bayesian networks; health-related problems; clinical outcomes; systematic review.

Resumo

Objetivo: O objetivo deste trabalho foi analisar sistematicamente a literatura sobre redes Bayesianas temporais e a sua relação com *outcomes* clínicos.

Introdução: O conhecimento médico tem conhecido enorme expansão. Apesar da maior facilidade de obtenção de dados pertinentes tornar a tarefa de diagnóstico mais eficaz, é também progressivamente mais complexa a adaptação por parte do médico. A elevada variedade de estudos sobre redes Bayesianas temporais e as suas aplicações na literatura, justifica a realização de uma revisão sistemática relativa à aplicação desta técnica na área da saúde.

Métodos: Esta revisão sistemática foi realizada em concordância com as *guidelines* do PRISMA. A pesquisa foi efetuada em várias bases de dados *online*, como a PubMed, ISI Web of Knowledge, SCOPUS e DBLP, sem restrições de data ou desenho de estudo. Foi aplicada metodologia de revisão sistemática padrão para a identificação de estudos elegíveis, extração de dados e avaliação do risco de viés.

Resultados: Em 8761 estudos, 42 foram considerados elegíveis para inclusão, tendo sido reunida evidência sobre redes Bayesianas temporais. Vinte e três (55%) estudos abordaram redes Bayesianas dinâmicas, dezasseis (38%) versaram o sistema cardiovascular e vinte e oito (67%) foram derivados para prognóstico. A análise demonstrou que a maioria dos estudos encontrou benefícios no uso de redes Bayesianas temporais. Foi encontrada elevada heterogeneidade entre os estudos, impossibilitando a realização de meta-análise.

Discussão: Este estudo tinha a ambição de realizar uma revisão sistemática, descobrindo novos dados e conteúdo sobre redes Bayesianas temporais aplicadas na área da saúde. Existe um considerável número de estudos que investigam os benefícios do uso deste tipo de redes em problemas relacionados com a saúde. Estes resultados podem incentivar o uso de redes Bayesianas temporais, dado o aumento de performance preditiva destas redes. A evidência e/ou modelos resultantes podem ser usados no futuro desenvolvimento de sistemas de suporte à decisão clínica.

Palavras-chave: redes bayesianas temporais; problemas na saúde; outcomes clínicos; revisão sistemática.

Preamble

As a master's student with healthcare background, the present study was motivated by the constant growing of health-related data and the need to facilitate both the management and interpretation of the same.

That way rises the interest on Temporal Bayesian Networks, which brings a different display mode, being capable to deal with multiple factors at multiple time points.

The conclusion of this study got one year delayed for professional reasons.

This study was invited for an apresentation at the Annual Symposium of the master's degree in Medical Informatics of the Faculty of Medicine, University of Porto (oral presentation on September 23th, 2017).

Table of Contents

Tables

Figures

Acronyms

AUC: area under the curve;

BMI: body mass index;

BN: bayesian network;

BNC: bayesian network classifier;

BS: brier score;

CAD: coronary artery disease;

CHD: coronary heart disease;

COPD: chronic obstructive pulmonary disease;

CVD: cardiovascular diseases;

CTBN: continuous time bayesian network;

DBBN: dynamic bayesian belief network;

DBN: dynamic bayesian network;

DBNC: dynamic bayesian network classifier;

DDO-DBN: data-driven only dynamic bayesian network;

DID: dynamic influence diagrams;

DSBN: dynamic stage bayesian network;

EI-DBN: expert intervention dynamic bayesian network;

FPR: false positive rate;

HDP: heart disease program;

IBD: inflammatory bowel disease;

ICU: intensive care unit;

ILI: influenza like illness;

ITBN: irregular time bayesian network;

LASSO: least absolute shrinkage and selection operator;

LBR: lazy bayes rule;

MDL: minimal description length;

MLR: multilevel regression;

mTBN: modifiable temporal bayesian network;

MTBN: multilevel temporal bayesian network;

NPEDT: networks of probabilistic events in discrete time;

NPV: negative predictive value;

PALF: pediatric acute liver failure;

PBN: prognostic bayesian network;

POMDP: partially observable markov decision processes;

PPV: positive predictive value;

PRISMA: preferred reporting items for systematic reviews and meta-analyses;

PTN: probabilistic temporal network;

P&I: pneumonia or influenza;

ROC: receiver operating characteristic;

SOFA: sequential organ failure assessment;

SMOTE-N: synthetic minority oversampling technique for nominal features;

SVM-RFE: support vector machine recursive feature;

TA: temporal abstraction;

TANB: temporal augmented naïve bayes;

TBN: temporal bayesian network;

TBNE: temporal bayesian network of events;

TNBN: temporal nodes bayesian networks;

TPR: true positive rate;

TRL: technology readiness level

VAP: ventilator-associated pneumonia;

1. Introduction

Medical knowledge is growing at a fiery rate. While the obtainability of pertinent data has the potential to make the task of diagnosis more accurate, it is also progressively overwhelming for medical staff to adapt [1].

Heijden et al. stated that medical data frequently appears as time series, offering a description of the development of a disease over time. Awareness into the evolution of a disease is a major aspect of the management of any disease, whether acute or chronic, for chronic patients the evolution is even more important, given that the disease will not disappear. For chronic diseases it is clinically interesting to study the relation between the different variables, approaching signs and symptoms, environmental factors, in terms of equally static and temporal relationships [2]. Bayesian networks (BN), introduced in the 1980s, where implemented to represent and reason with models of problems regarding issues like uncertainty, adopting, as a basic framework, the probability theory [3].

During the last decade, due to their capability to model complex systems, where such relationships between the many variables involved are not entirely known, BN have raised much interest in medicine. Also, they offer a causal interpretation instead of merely capture association. Given that, they have become popular as prognostic models [4]. BN are increasingly used in biomedicine and health-care to support different types of problem interpreting, such as diagnostic reasoning, prognostic reasoning, and treatment selection and learning functional interactions [5].

In a study carried out by Visscher et al. it was found improved performance in the prediction of pathogens causing ventilator-associated pneumonia using a BN model [6]. According to Young, BN is a strong probabilistic technique of dealing with uncertainty, which became popular between artificial intelligence researchers. Nevertheless, one should be aware that BN do not directly provide a mechanism to represent temporal dependencies [7].

In medicine, for many tasks such as prevention, diagnosis, therapeutic management, prognosis and discovery, representing and reasoning about time is crucial [8]. According to Figueroa, BN are one of the most used techniques for the development of intelligent systems, and although BN were not intended to model temporal aspects explicitly, recently BN have been applied to reasoning about time under uncertainty [8].

Temporal Bayesian Networks (TBN) are temporal extensions of BNs, explicitly representing probabilistic relationships between variables, in the form of graphical models. Since they have the capability to represent and reason knowledge under uncertainty, TBN are highly applicable in the medical field. TBN are particularly important to deal with the explicit or implicit representation of time. That way, they have many applications in medicine, like medical diagnosis, forecasting, and medical decision making. Also, they have been used in many clinical domains to interpret and reason with several sums of longitudinal data [9].

According to the literature, there are many types of TBN, although the most popular temporal extension of BNs is dynamic Bayesian networks (DBN), which uses a discrete-time representation. Extensions of DBNs for decision making are dynamic influence diagrams (DID) and partially observable Markov decision processes (POMDP). Models using interval-based representation of time are: networks of probabilistic events in discrete time (NPEDT), modifiable temporal Bayesian networks (mTBN), probabilistic temporal networks (PTN) and temporal nodes Bayesian networks (TNBN). Also, models using continuous time representation are the continuous time Bayesian networks (CTBN) and Berzuini's network of date's model. Irregular time Bayesian networks (ITBN) are a new temporal extension of BNs which are capable to deal with processes happening irregularly through time [9].

Temporal probabilistic networks are now being used in many medical problems such as diagnosis, treatment selection, therapy monitoring and prognosis. Given their ability to deal with uncertainty in time-series medical data they permit one to learn about causal relationships and dependencies of clinical features [9]. There are many studies regarding TBN models and clinical data. For example, there is evidence of the use of temporal reasoning to improve the diagnosis accuracy, on a system called Heart Disease Program (HDP) [10]. Charitos et al., with the application of a DBN for the diagnosis of ventilator-associated pneumonia (VAP) in Intensive Care Unit (ICU) patients, found that representing time explicitly and taking into account the patient's history, rises diagnostic performance. In the same study, the DBN shown a better performance at distinguishing between VAP and non-VAP patients than the static network [11]. Another example is the usage of a DBN model in the diagnosis of apnea [12]. DBN were also used to describe changes in organ failures in ICU patients [4].

Figueroa also talks about Temporal Bayesian Networks of Events (TBNE) in which each node represents an event or state change of a variable, and an arc corresponds to a causal-temporal relationship. A temporal node represents the time a variable changes is state, including an option of no-change. The temporal intervals can vary in number and size for each temporal node, allowing multiple granularity. Temporal information is relative, that is, there is not an absolute reference about time. A TBNE creates a formal and systematic structure, modelling the temporal evolution of dynamic process [8]. Lappenschaar gave an additional contribute developing the concept of Multilevel Temporal Bayesian Network (MTBN), with the capability of analyzing multiple outcomes at multiple time points [13].

Given the exposed, the wide range of TBN models and applications available in the literature, a formal systematic review of TBN applications to medical problems was needed.

The aim of our study was to systematically review the literature on Temporal Bayesian Networks (TBN) and its relation with clinical outcomes. Thus, this systematic review will answer the following question:

> *Are there significant benefits on the use of Temporal Bayesian Networks on health-related problems when compared to traditional Bayesian networks or other statistical methods?*

2. Aim

The objective of this study, as stated above, was to perform a systematic review of the available literature on TBN and its relations with clinical outcomes, attempting to gather information about benefits or disadvantages of this technique, when compared to other types of techniques.

For that we defined the following specific objectives:

- To systematically review the literature on TBN and its relation with clinical outcomes;
- To identify the benefits of TBN's application on health-related outcomes;
- To discover data that might be used in the future development of clinical decision support systems.

3. Material and Methods

This study is a systematic review that ran from January 2016 to September 2017, with the last search being done in February 2016. The methodology of this review is reported according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) group[14].

3.1. Eligibility criteria

We aggregated evidence about studies that address TBN and their applicability to health research and clinical outcomes.

3.2. Information sources

Search was conducted on multiple online databases, including PubMed, ISI Web of Knowledge, SCOPUS and DBLP.

3.3. Search strategy

There was no study design or date limits on the search. Platforms mentioned above were searched using specific queries built with proper keywords and, when relevant, classification terms (e.g. MeSH). The proposed query below was adapted for each database:

((bayes OR belief OR Bayes Theorem [Mesh] OR causal*) AND (network*)) AND (("Longitudinal Studies"[Mesh]) OR "Spatio-Temporal Analysis"[Mesh] OR temporal OR continuous OR dynamic* OR longitudinal)*

3.4. Data management

In this study, all searches were performed independently. All studies were aggregated on a specific software, to avoid eventual duplicates. The studies were properly identified, univocally, by a specific identifier. The result of the review process was exported to a data analysis software.

3.5. Selection process

All studies were independently reviewed by two reviewers, both in the initial phase of titles and abstracts, and in the full-text review phase, with the application of the defined inclusion criteria. Disagreement was solved through discussion and consensus.

3.6. Data collection process

For the application of the eligibility criteria a specific form was used, containing the following information.

- Study ID;
- Inclusion Criteria:
    - Reference to Temporal Bayesian Networks (the study must approach model(s) of Temporal Bayesian Networks);
    - Reference to Health (the study must approach health-related problems).

Similarly, for the studies' data extraction, an appropriate form was applied, whose items are described in the following area.

3.7. Data items

In this review, we have extracted the following data items:

- Study ID;
- Type of applied Bayesian network (this section contains the type of Bayesian network applied, for instance, Dynamic Bayesian Networks);
- Compared technique (in this section it was registered the model compared with the Bayesian network model, such as multilevel regression);
- Medical specialty (in this topic it was stated the medical specialty in study, such as Cardiology, Neurology, etc);
- Clinical area (this topic comprised the area of study, such as diagnosis, prognosis, therapy, etc);

- Modeling goal (this sector approached the purpose of the modelling such as improving the diagnosis for a determined pathology);
- Model evaluation measure (in this section it was evaluated the model's measures used by the authors, such as sensitivity, specificity, the accuracy of a determined prediction, etc);
- Data collection methods (this area approached the data collection methods, such as forms);
- Model validation techniques (this section specified the data model validation techniques used by the authors, such as hold-out technique, cross-validation, etc);
- Cohort description (this topic approached the size of the sample);
- Development maturity (whether the study reported a model, a pilot study, a prototype or a final product);
- Country;
- Year.

3.8. Outcomes

This review included studies that report the following outcome measures:

- Model's applicability;
- Model's benefits/advantages comparatively with other techniques;
- Applications maturation state.

3.9. Risk of bias assessment

The risk of bias assessment was conducted in accordance with QUADAS-2[15].

3.10. Meta-analysis

Since there was high heterogeneity of studies no meta-analysis was conducted regarding dealing with missing data, data synthesis, subgroup analysis and investigation of heterogeneity or assessment of publication bias.

3.11. Ethics approval and consent to participate

Not applicable.

3.12. PROSPERO registration

Although we have tried to register our protocol with PROSPERO, this has been seen as not required since *"this is a methodology review which has no directly health related outcomes."* PROSPERO's Committee answer follows attached as Annex 1.

4. Results

After applying all of the above criteria, 42 studies were selected, as shown in Figure 1. A first impression on the selected papers allowed us to place them in terms of type, time, and location while exploring the issues each of them covers.

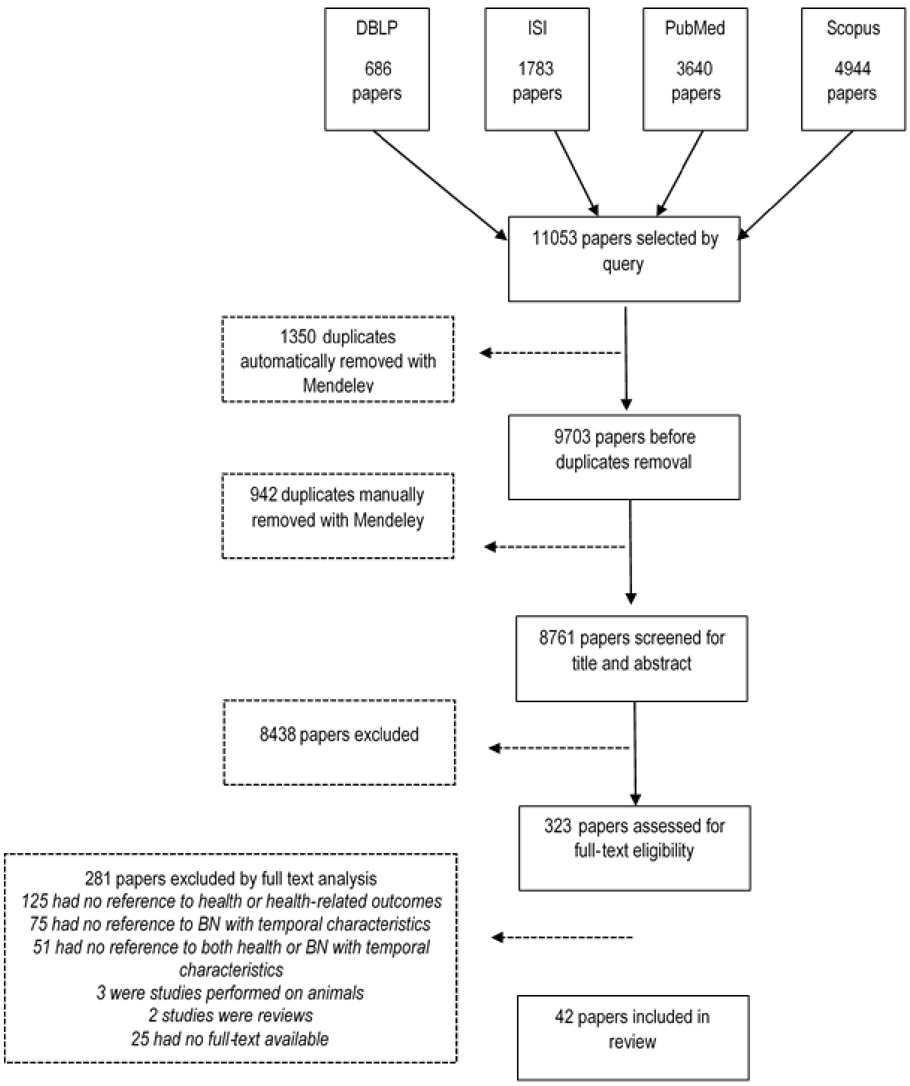

Figure 1. Study Selection

4.1. Descriptive analysis

We observed that 69% of the articles were originated from Europe, 28% of these were originated from Netherlands and 21% of the papers came from the United States of America. The great majority of the articles were published between 2012 and 2015 (55%) being 2014 the year of bigger contribution with 8 papers. In the field of development maturity technology readiness level (TRL) was used. We categorized 15 studies as TRL1 (36%), 24 studies as TRL2 (57%), 2 as TRL3 (5%) and 1 as TRL4 (2%). Regarding the applications maturation state 24 were categorized as needed to be tested (57%) and 18 as ready to use (43%). Their main characteristics are summarized in Table 1.

Risk of bias assessment results are exposed in Table 2. It was possible to find low risk of bias in twenty three studies (55%) concerning patient selection; regarding index test, forty studies (95%) were considered as of low risk of bias; reference standard risk of bias was also assessed with twenty eight studies (66%) founded as unclear for risk of bias with thirty four studies (81%) presenting low risk of bias in terms of flow and timing. About applicability concerns, in patient selection's category we found thirty one studies (74%) with low risk of introducing bias; concerning index test's applicability it was possible to find thirty nine studies (93%) with low risk; at last the reference standard's applicability was found to have more risk of introducing bias, with twenty eight studies (66%) presenting high risk of bias.

4.1.1. Type of Applied Bayesian Network

In this section we will expose the different types of BNs applied in the studies. It is important to notice that during the study, the most common type of BN was DBN which, in this study, is considered a type of TBN.

Twenty three of forty two studies addressed DBN (55%)[4, 11, 18, 39, 27, 28, 35, 30, 17, 16, 23, 29, 49, 45, 47, 36, 19, 20, 34, 37, 21, 26, 46]; six of these twenty three studies dedicated their efforts to Cardiovascular diseases (CVD), approaching themes like hypertension[18], electrocardiography for myocardial infarction classification[39], prediction of coronary atherosclerosis[27, 28], prediction of survival in patients with ischaemic heart disease[35] and in patients with ventricular assist devices[30]; it was also possible to find six studies addressing different areas of Oncology[17, 16, 23, 29, 49, 45]; the Trauma, emergency room and ICU medical specialty was found in five out of twenty five studies[4, 47, 36, 19, 20]; the remaining studies varied in their medical specialty application, from Pulmonology[11], Diabetes[34, 37], to the category of Others, addressing probability of death and spontaneous survival in pediatric acute liver failure patients[21], risk of fall from gait[26] and hemodialysis[46].

Study	Location	Dataset	Applied BN	Comparison	Medical Specialty	Clinical area
Akutekwe (2014)[16]	Europe	16331 features (112 patients)	DBN	N.A.	Oncology	Diagnosis
Akutekwe (2014)[17]	Europe	592 features (72 patients)	DBN	N.A.	Oncology	Prognosis
Akutekwe (2014)[18]	Europe	22184 predictor genes (159 patients)	DBN	N.A.	CVD	Diagnosis
Almahmoud (2015)[19]	North America	472 patients	DBN	N.A.	Trauma	Prognosis
Almahmoud (2015)[20]	North America	472 patients	DBN	N.A.	Trauma	Prognosis
Azhar (2013)[21]	North America	216 serum samples (49 patients)	DBN	N.A.	Others	Prognosis
Ceccon (2011)[22]	Europe	62 patients	DSBN	DBN/BN	Ophthalmology	Prognosis
Chang (2008)[23]	Germany	231 human breast cancer cells	DBN	N.A.	Oncology	Diagnosis
Charitos (2009)[11]	Europe	157 patients	DBN	Static network	Pulmonology	Diagnosis
Chiang (2010)[24]	Asia	905 heartbeats	DBN with CBFM	N.A.	CVD	Diagnosis
Codecasa (2014)[25]	Europe	120 trajectories for exercise	CTBN	DBN	CVD	Prognosis and rehabilitation
Cuaya (2013)[26]	North America	66 records (18 patients)	DBN	N.A.	Others	Prognosis
Exarchos (2013)[27]	Europe	39 patients	DBN	N.A.	CVD	Prognosis
Exarchos (2015)[28]	Europe	39 patients	DBN	N.A.	CVD	Prognosis
Exarchos (2012)[29]	Europe	86 patients	DBN	N.A.	Oncology	Prognosis
Exarchos (2014)[30]	Europe	71 patients	DBN	N.A.	CVD	Prognosis
Gatti (2012)[31]	Europe	3 case studies	CTBN	DBN	CVD	Diagnosis and prognosis
Kayaalp (2000)[32]	North America	1449 patients	STBN	NSTBN	Trauma	Prognosis
Lappenschaar (2013)[13]	Europe	1.5M patients	MTBN	N.A.	CVD	Prognosis
Lappenschaar (2013)[33]	Europe	218333 patients	MBN	MLR	CVD	Prognosis
Marini (2015)[34]	Europe	1441 patients	DBN	N.A.	Diabetes	Prognosis
Marshall (2010)[35]	Europe	2512 patients	DBN	N.A.	CVD	Prognosis
Nachimuthu (2012)[36]	North America	3100 patients	DBN	N.A.	Trauma	Diagnosis
Nachimuthu (2010)[37]	North America	796 patients	DBN	Rule-based protocol	Diabetes	Prognosis
Olesen (2006)[38]	Europe	2414 questionnaires	TPSBN	BN	Others	Diagnosis
Oliveira (2011)[39]	South America	Not specified	DBN	BN	CVD	Diagnosis
Orphanou (2012)[40]	Europe	176 patients	DBN-TA	N.A.	CVD	Prognosis
Orphanou (2014)[41]	Europe	1428 patients	DBN-TA	N.A	CVD	Diagnosis
Orphanou (2015)[42]	Europe	Not specified	DBN-TA	N.A.	CVD	Diagnosis
Orphanou (2015)[43]	Europe	1417 patients	DBN-TA	DBN	CVD	Prognosis
Peelen (2010)[44]	Europe	2271 patients	HDBN	N.A.	Trauma	Prognosis
Peeling (2007)[45]	Europe	26+162 patients	DBN	BN	Oncology	Diagnosis and Prognosis
Rose (2005)[46]	Asia	Not specified	DBN	N.A.	Others	Prognosis
Sandri (2014)[4]	Europe	79 patients	DBN	N.A.	Trauma	Prognosis
Sebastiani (2005)[47]	North America	Not specified	DBN	N.A.	Trauma	Prognosis
Tucker (2004)[48]	Europe	180+24 patients	STBN	BN/Linear regression/k-nn	Ophthalmology	Diagnosis
Van der Heijden (2014)[2]	Europe	10 patients	TANB/TNBN	Aerial/Bootstrap/Expert	Pulmonology	Prognosis
Van Gerven (2008)[49]	Europe	3 case studies	DBN	N.A.	Oncology	Prognosis
Verduijn (2007)[50]	Europe	10147 patients	PBN	MDL	CVD	Prognosis
Visscher (2007)[51]	Europe	2424 patients	NRDBM	N.A.	Pulmonology	Prognosis
Watt (2011)[52]	North America	50+4700 patients	DBBN	LBR-based model	Oncology	Prognosis
Watt (2008)[53]	North America	4796 patients	DBBN	Logistic regression	Others	Prognosis

Table 1. Characteristics of the included studies

Study	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Akutekwe (2014)[16]	(-)	(+)	(?)	(-)	(?)	(+)	(-)
Akutekwe (2014)[17]	(-)	(+)	(?)	(-)	(?)	(+)	(-)
Akutekwe (2014)[18]	(-)	(+)	(?)	(-)	(?)	(+)	(-)
Almahmoud (2015)[19]	(-)	(+)	(?)	(+)	(?)	(+)	(-)
Almahmoud (2015)[20]	(-)	(+)	(?)	(+)	(?)	(+)	(-)
Azhar (2013)[21]	(+)	(+)	(?)	(-)	(+)	(+)	(-)
Ceccon (2011)[22]	(-)	(+)	(+)	(+)	(?)	(+)	(+)
Chang (2008)[23]	(?*)	(+)	(?)	(+)	(?)	(+)	(-)
Charitos (2009)[11]	(+)	(+)	(+)	(+)	(+)	(+)	(-)
Chiang (2010)[24]	(+)	(+)	(?)	(?)	(+)	(+)	(-)
Codecasa (2014)[25]	(+)	(+)	(+)	(+)	(+)	(+)	(+)
Cuaya (2013)[26]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Exarchos (2013)[27]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Exarchos (2015)[28]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Exarchos (2012)[29]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Exarchos (2014)[30]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Gatti (2012)[31]	(?**)	(+)	(+)	(+)	(-)	(+)	(+)
Kayaalp (2000)[32]	(+)	(+)	(-)	(+)	(+)	(+)	(+)
Lappenschaar (2013)[13]	(-)	(+)	(?)	(-)	(+)	(+)	(-)
Lappenschaar (2013)[33]	(?***)	(+)	(+)	(+)	(+)	(+)	(+)
Marini (2015)[34]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Marshall (2010)[35]	(-)	(+)	(?)	(-)	(+)	(+)	(-)
Nachimuthu (2012)[36]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Nachimuthu (2010)[37]	(+)	(+)	(+)	(+)	(+)	(-)	(+)
Olesen (2006)[38]	(?****)	(+)	(?*****)	(+)	(?****)	(+)	(?*****)
Oliveira (2011)[39]	(-)	(+)	(+)	(+)	(?)	(+)	(+)
Orphanou (2012)[40]	(+)	(+)	(?)	(+)	(+)	(-)	(-)
Orphanou (2014)[41]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Orphanou (2015)[42]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Orphanou (2015)[43]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Peelen (2010)[44]	(+)	(+)	(?)	(+)	(+)	(+)	(-)
Peeling (2007)[45]	(?******)	(+)	(+)	(+)	(?******)	(+)	(+)
Rose (2005)[46]	(-)	(+)	(?)	(+)	(+)	(+)	(-)
Sandri (2014)[4]	(+)	(+)	(?)	(+)	(+)	(-)	(-)
Sebastiani (2005)[47]	(+)	(+)	(?)	(-)	(+)	(+)	(-)
Tucker (2004)[48]	(+)	(+)	(+)	(+)	(+)	(+)	(+)
Van der Heijden (2014)[2]	(-)	(+)	(+)	(+)	(+)	(+)	(+)
Van Gerven (2008)[49]	(-)	(+)	(?)	(+)	(+)	(+)	(-)
Verduijn (2007)[50]	(+)	(-)	(-)	(+)	(+)	(+)	(+)
Visscher (2007)[51]	(-)	(+)	(?)	(+)	(+)	(+)	(-)
Watt (2011)[52]	(-)	(+)	(+)	(+)	(+)	(+)	(+)
Watt (2008)[53]	(+)	(-)	(-)	(+)	(+)	(+)	(+)

Table 2. Risk of bias of the included studies

(+) = low risk; (-) = high risk; (?) = unclear risk.
*synthetic dataset
**no dataset – scenario analysis
***use of both real and synthetic dataset
****dataset based on a survey from another survey
*****compared to other study
******dataset from cross-section studies

Four of forty two studies approached DBN with temporal abstraction (TA) (9%) in order to predict the progression and estimate the risk of coronary heart disease[40, 42, 41, 43]. Two of forty two studies (5%) focused on post-stroke rehabilitation[25] and prediction of complications and persistence of cardiogenic heart failure[31] using CTBN. It was also possible to discover two studies (5%) that applied the concept of Dynamic Bayesian Belief networks (DBBN) addressing neuro-oncology[52] and osteoarthritic knee pain[53].

The remaining eleven studies (26%) approached different types of BNs, such as Dynamic Bayesian Network with Constraint-based fuzzy models[24], MTBN[13], Multilevel Bayesian Network[33], Prognostic Bayesian Network (PBN)[50], Hierarchical Dynamic Bayesian Network[44], Dynamic Stage Bayesian Network (DSBN)[22], Spatiotemporal Bayesian Network[48], Stationary Temporal Bayesian Network[32], Non-repetitive Dynamic Bayesian Multinetwork[51], TNBN[2] and Tailored Patient Specific Bayesian Network[38].

This information is summarized in Table 3:

	DBN	**DBN with TA**	**Continuous Time BN**	**DBBN**	**Other type of BN**
Medical Specialty (n=42)	CVD[18, 39, 27, 28, 35, 30]	CVD [40, 42, 41, 43]	CVD[25, 31]	Oncology[52]	CVD[13, 24, 33, 50]
	Oncology[17, 16, 23, 29, 49, 45]			Others[53]	Trauma, emergency room and ICU[44, 32]
	Trauma, emergency room and ICU[4, 47, 36, 19, 20]				Pulmonology[2, 51]
	Pulmonology[11]				Ophthalmology[22, 48]
	Diabetes[34, 37]				Others[38]
	Others[21, 26, 46]				

Table 3. BN distribution by medical specialty

4.1.2. Medical Specialty

In this section we will expose the different medical specialties of the studies.

Sixteen of forty two studies addressed the CVD (38%)[13, 18, 39, 27, 28, 35, 30, 40, 42, 41, 43, 25, 31, 24, 33, 50]; most of them dedicated their efforts to predict the progression and estimate the risk of coronary heart disease[27, 28, 40, 42, 41, 43] while others focused on survival prediction[35] and interaction with other diseases[33] of heart failure. Although some of the studies focused on disease progression, mortality, multimorbidity, some were derived for diagnosis[18, 39, 42, 41, 24], others for prognosis[13, 27, 28, 35, 30, 40, 43, 33, 50], one for both diagnosis and prognosis[31] and one for prognosis and rehabilitation[25].

Regarding the medical specialty of Oncology we got seven of forty two studies (17%)[17, 16, 23, 29, 49, 45, 52]. The great majority of these studies approached disease's prognosis[17, 29, 49, 52], while other focused on diagnosis[16, 23] and one of the studies addressed both diagnosis and prognosis[45].

Concerning the field of Trauma, emergency room and ICU we were able to find seven studies (17%)[4, 47, 36, 19, 20, 44, 32]. The great majority of them were derived for disease's prognosis[4, 47, 19, 20, 44, 32], only one addressed diagnosis[36].

In the Pulmonology medical specialty we discovered three articles (6%)[2, 11, 51]. Two of them approached the risk of developing ventilator-associated pneumonia[11, 51] and the other focused on the prediction of exacerbations in subjects with COPD[2].

There were two studies for each of these medical specialties: Ophthalmology (5%)[22, 48] and Diabetes (5%)[34, 37]. Three studies were derived for prognosis, addressing visual field deterioration in glaucomatous subjects[48], approaching type-1 diabetes complications[34], the prediction of serum glucose levels of ICU patients[37]; with one study approaching the modelling key stages in the development of glaucoma[22] inserted in the diagnosis category.

At last, four articles approaching hemodialysis, knee pain, pediatric acute liver failure and risk of fall from gait, categorized as Others (12%), were all derived for prognosis[21, 26, 46, 53], with only one, regarding immunology that was derived for diagnosis[38].

This information is summarized in Table 4:

	CVD	Oncology	Trauma/Emergency Room/ICU	Pulmonology	Ophthalmology	Diabetes	Others
Clinical Area (n=42)	Prognosis[13, 27, 28, 35, 30, 40, 43, 33, 50]	Prognosis[17, 29, 49, 52]	Prognosis[4, 47, 19, 20, 44, 32]	Prognosis[11, 51]	Prognosis[48]	Prognosis[34, 37]	Prognosis[21, 26, 46, 53]
	Diagnosis[18, 39, 42, 41, 24]	Diagnosis[16, 23]	Diagnosis[36]	Diagnosis[2]	Diagnosis[22]		Diagnosis[38]
	Prognosis and Diagnosis[31]	Prognosis and Diagnosis[45]					
	Prognosis and Rehabilitation[25]						

Table 4. Medical specialty distribution by clinical area

4.1.3. Clinical Area

In this section we will expose the different clinical areas of the studies.

Twenty eight of forty two studies were derived to prognosis (67%), those studies addressed different medical specialties like CVD[13, 27, 28, 35, 30, 40, 43, 33, 50], Oncology[17, 29, 49, 52], Trauma, emergency room and ICU[4, 47, 19, 20, 44, 32], Pulmonology[2, 51], Ophthalmology[22], Diabetes[34, 37] and the section categorized as Others[21, 26, 46, 53].

Regarding the clinical area of diagnosis we got eleven of forty two studies (26%). The great majority of these studies addressed cardiovascular system[18, 39, 42, 41, 24], two of eleven studies addressed the oncology specialty[16, 23]. The remaining studies were divided in medical specialties like Trauma, emergency room and ICU[36], Pulmonology[11], Ophthalmology[48] and Others[38].

The remaining studies are gathered in the clinical areas of diagnosis and prognosis (5%), and prognosis/rehabilitation (2%). diagnosis and prognosis contemplates two studies, one addressing cardiogenic heart failure[31] and another addressing B-cell lymphoma and glaucoma[45]. Prognosis/rehabilitation considers a study about post-stroke rehabilitation[25].

Table 5 shows the distribution of the clinical areas by the type of applied BN:

	Prognosis	Diagnosis	Diagnosis and Prognosis	Prognosis and Rehabilitation
Type of BN	DBN[4, 27, 28, 35, 30, 17, 29, 49, 47, 19, 20, 34, 37, 21, 26, 46]	DBN[11, 18, 39, 16, 23, 36]	DBN[45]	CTBN[25]
	DBN with TA[40, 43]	DBN with TA[42, 41]	CTBN[31]	
	DBBN[52, 53]	Others[24, 48, 38]		
	Others[2, 13, 33, 50, 44, 22, 32, 51]			

Table 5 - Clinical area distribution by BN

4.2. Applications of Dynamic Bayesian Networks

This section will point out the results obtained by the twenty five articles addressing DBN, splitting each sub-section by medical specialty.

4.2.1. Cardiovascular Diseases

In this systematic review it was possible to found six articles regarding both DBN and CVD. Exarchos et al. carried out two studies approaching coronary atherosclerosis, in order to identify key factors on the progression of the disease. In one of these studies[27] the model achieved an overall accuracy of 79%, finding also 5 features that affect more prominently the progression of the disease, including diabetes, cholesterol and family history of coronary disease; the other study[28] also tried to identify the factors that play a key role towards the progression of atherosclerosis, employing these factors in a decision support system in order to conjecture about the prognosis of the disease. The results were divided in a baseline and a temporal analysis. The temporal analysis outperformed the baseline analysis both in sensitivity, specificity and accuracy, for the two major outcomes – number of stenosis and atherosclerosis progression. Akutekwe and Seker[18] applied a DBN using five feature selection methods, allowing the discovery of possible key biomarkers for the diagnosis of hypertension, with least absolute shrinkage and selection operator (LASSO) and support vector machine recursive feature (SVM-RFE) linear methods showing the best predictive performance. Oliveira et al.[39] performed a comparison between a DBN and a BN, regarding the use of a network for the detection of cardiac arrhythmias, with the DBN outperforming the BN with an improvement of 6% in terms of sensitivity. Marshall et al.[35] tried to establish an association between patient variables, the risk of developing CVD and survival, with is main findings being the influence of diastolic blood pressure on survival at 5 years follow-up period, also the body mass index (BMI) seems to have an impact on survival. Again Exarchos et al.[30] developed a decision support tool to predict the survival probabilities of patients after the implantation of a ventricular assist device. Two DBN were developed, a short term and a long term, with both networks reaching values of accuracy of 96,97% and 93,55% respectively.

4.2.2. Oncology

This section will approach the studies where it was possible to find both DBN and oncologic diseases.

Akutekwe and Seker contributed to this area with two studies. One of the studies[17] applied four feature selection methods, allowing the discovery of a possible temporal association of high quality metabolic features for ovarian cancer, with LASSO showing the best predictive accuracy (93%). Two DBN were used to model the temporal relationships among 39 features, showing significant relationships among features 219 and 225 and between features 534 and 219 through time points. It was also found 20 metabolic chemical compounds that can be associated to ovarian cancer. The same authors also applied[16] a DBN using feature selection methods, thus allowing the discovery of a possible temporal association of high quality biomarker subsets for colorectal cancer diagnosis, but it was not possible to found objective outcomes on this study, the same was evaluated in a subjective way. Gerven et al.[49] developed a prognostic model and applied it in three low-grade carcinoid patients taking into account the prediction of the disease evolution, but like in the previous study there was no presentation of model evaluation measures. Peeling and Tucker[45] approached two datasets (B-cell lymphoma microarray data and glaucoma visual field data) using two types of classifiers. Bayesian network classifier (BNC) learnt with no temporal information and dynamic bayesian network classifier (DBNC) learnt from pseudo time-series, in order to check the relevance of temporal relationships on the classification performance. In the b-cell lymphoma dataset, with a ordered dataset on the dataset, the DBNC was the best with an improved area under the curve (AUC). Although, a test was performed using random ordering and the BNC outperformed the DBNC with sensitivity of 0.70 and AUC of 0.84. Regarding the glaucoma dataset, the authors found and improvement in the AUC when using the DBNC. In a study carried out by Chang et al.[23] it was performed quantitative bayesian inference and model averaging using a set of qualitative statements. This approach enabled probabilistic inference by model averaging allowing the prediction of probabilistic quantities from a set of qualitative constraints. A use case regarding breast cancer is performed, showing realistic quantitative predictions by prior knowledge alone, although when additional information about ranges of gene expression is taken into account the values get closest to the true values. At last Exarchos et al.[29] created a DBN to identify factors that dictate oral squamous cell carcinoma progression and to predict potential disease relapses. This study obtained accuracy results of 86%, with specificity and sensitivity values of 63,6% and 100% respectively towards relapse probability at first follow-up visit . When the authors aggregate baseline analysis and first follow-up visit towards the prediction of relapse probability at second follow-up visit, the results obtained are 100% in sensitivity, specificity and accuracy. Compared with previous literature this study has a bigger sample and also greater accuracy values (100%).

4.2.3. Trauma, emergency room and ICU

This section will focus on the studies that presented both DBN and Trauma, emergency room and ICU related diseases.

Almahmoud et al, gave their contribute with two studies, in one of the studies[20] the authors tried to analyze the impact of prehospital hypotension on systemic acute inflammatory response to blunt trauma, using a DBN to infer for inflammatory biomarkers within 24 hours post-injury, finding feedback effects of some biomarkers on their own productions and showing an effect in the inflammatory response in the hypotensive cohort, with the same happening on the normotensive group. The other study of the same authors[19] used a DBN to model the evolution of probabilistic dependencies within a system over time, in this case to understand the relationship between trauma and inflammatory mediators. The DBN findings correlate with those from prior studies, addressing the role of different chemokines as central mediators of systemic inflammation with a variation on the gravity of the injury. It was also possible to find associations between chemokines that is one chemokine level will affect the level of another. Regarding the presence of sepsis soon after emergency department arrival, Nachimuthu and Haug[36] constructed a DBN with 3,6, 12 and 24 timeslices for each patient. The results presented AUC of 0.911, 0.915, 0.937 and 0.944 respectively. It was also possible to found values of Sensitivity, Specificity, positive predictive value (PPV), negative predictive value (NPV) and F-measure. The respective values increase further as more data becomes available. A study performed by Sandri et al.[4] applied a DBN in order to model sequential organ failure assessment (SOFA) score changes in adult critically ill patients, aiming to identify the most probable sequence of organ failures in the first week after ICU admission in a three slices network. The constructed DBN presented accuracy ranging from 71.62% to 74.95%, that when compared to other studies are similar. Additionally, this work showed the existence of organ failure sequences more likely to occur than others. At last Sebastiani et al.[47] approached the symptoms and mortality of patients with influenza and/or pneumonia divided by age in two groups, children and adults. It was found a good performance by the DBN. With the children data giving predictive insight to the adult data, influenza like illness (ILI) data and of the number of pneumonia or influenza (P&I) deaths. The opposite didn't occur, neither of the other datasets were predictive for the children data. This might be explained by the earlier and stronger signals of influenza in children.

4.2.4. Pulmonology

In the Pulmonology medical specialty it was possible to found one article considering both DBN and health-related outcomes.

With Charitos et al.[11]developing a DBN to capture the development of VAP through time. The DBN performance was evaluated focusing on the diagnostic prediction per day. The overall Brier score (BS) for the DBN was 0.2376, which outperformed the original network that had a 0.3370 overall Brier score. The DBN also proved to distinguish better between VAP and non-VAP patients. However the original model outperformed the DBN when it came to diagnosis patients without VAP.

4.2.5. Diabetes

Regarding Diabetes and DBN it's possible to discover two papers. One of the studies was carried out by Marini et al.[34] where the authors attempted to simulate patient health state and complications through discretized variables, using two types of models, one entirely learned from the data, called data-driven only network (DDO-DBN) and other partially driven by literature derived knowledge called expert intervention (EI-DBN). Both models were evaluated by state error test and difference between tests (Kolmogorov-Smirnov). When compared to an edgeless network both models show a very low error for the whole period, with less than 10% in error state. When compared with each other they don't differ much, the average error obtained by the DDO-DBN network is only 0.3% lower than the one of the EI-DBN network. The other study carried out by Nachimuthu et al.[37] developed a DBN model to represent clinical variables and the temporal and attemporal relationships underlying insulin and glucose homeostasis. The developed model was compared to an existing rule-based protocol (e-Protocol Insulin), showing comparable accuracy. In some cases the DBN model recommended lower dosage of insulin, being the safer protocol, but the opposite also happened in some ranges of patients.

4.2.6. Other medical specialties

This section will contain three studies with each one approaching a different medical specialty by use of a DBN. Azhar et al.[21] focused on the development of a DBN in order to establish relationships between circulating inflammatory mediators as a function of time and pediatric acute liver failure (PALF), in an attempt to identify a possible mechanism of the disease. The DBN model allowed to distinguish between cytokines networks associated with survival and cytokines networks associated with non-survival. The authors concluded that DBN-defined inflammatory networks may be used as new biomarkers for predicting outcomes in PALF. Cuaya et al.[26] presented two DBN models to estimate the risk of falls in elderly using spatio-temporal gait data and further explore the use of feature selection algorithms to automatically choose the gait parameters relevant to predict falls. One model was built using gait variables and the other one was constructed using gait variables picked only by feature selection algorithm. Both models showed average precision of 70% in predicting both imminent falls and 6 months risk of falling, with some superiority from the model built using gait variables with 72.22% for both outcomes. The second model achieved 72.22% and 66.66% precision values in predicting both imminent falls and 6 months risk of falling, respectively. Both models presented good values of specificity and sensitivity. At last, Rose et al.[46] carried out a study about a decision support system for surveillance of patients suffering from renal failure and treated by hemodialysis. It was possible to notice the use of a DBN adapted to the monitoring of the dry weight of patients suffering from chronic renal failure treated by hemodialysis. The model was only subjective evaluated, with encouraging results according to both nephrologists and patients.

4.3. Applications of Dynamic Bayesian Networks with Temporal Abstraction

In the present section all studies address the CVD medical's specialty.

On this section we will approach the use of DBN with TA in the medical specialty of CVD. All of the studies had the same major investigator. In one of these studies Orphanou et al.[40] created a DBN with TA incorporation to make predictions for the risk of coronary artery disease (CAD) for a specific patient, making only a preliminary approach. In order to perform primary coronary heart disease (CHD) risk assessment using the STULONG dataset, Orphanou et al.[42] constructed a DBN with TA integration. The model achieved satisfactory values of Specificity, Sensitivity, Precision and F1score with 0.73, 0.75, 0.74 and 0.74 respectively, after clustering oversampling in the dataset. In another study[41] the same authors constructed a DBN with TA integration to perform CHD diagnosis on a particular patient based on the medical history. The authors divided the dataset in 4 different datasets, D1 (original dataset), D2 (defined by random oversampling the minority class), D3 (oversampling using synthetic minority oversampling technique for nominal features (SMOTE-N)), D4 (combination of oversampling with SMOTE-N and random undersampling) and developed four networks, one for each dataset. The worst classified dataset was D1 and D4 presented the best classification performance with 91%, 75% and 82% for Recall, Precision and F1score, respectively. The author also used ROC curves to evaluate the models with D4 dataset coming as the best classified. Again, the same authors presented[43] an extended DBN model that incorporates TA methods applied to the prognosis of the risk of coronary artery disease in the STULONG dataset. Given the imbalanced training dataset, the authors divided the dataset in 5 different datasets, D1 (original dataset), D2 (defined by random oversampling the minority class), D3 (oversampling using SMOTE-N), D4 (combination of oversampling with SMOTE-N and random undersampling) and D5 (oversampling using random oversampling technique and clustering undersampling). Then they developed 5 DBN-extended networks, one for each dataset, and 5 DBN networks without TA. In order to evaluate the model's performance the adopted metrics were Precision, Recall and F1score. From both groups of DBN the network correspondent to the D5 dataset presented the greatest performance. When comparing the DBN-extended network with the DBN network without TA, the first one outperformed the last one with 0.72, 0.75 and 0.74 for Precision, Recall and F1score, respectively. When it comes to ROC curves, again the DBN-extended network outperformed the DBN network without TA, with AUC values of 0.78 and 0.60, respectively.

4.4. Applications of Continuous Time Bayesian Networks

Another variation of DBN are the CTBN which will be addressed in this section, regarding the cardiovascular medical specialty.

There are two studies approaching this technique on this systematic review. One of the studies[25] had the goal of analyzing the performance of CTBN classifiers on post-stroke rehabilitation. The models presented in the paper have their performance analyzed on both synthetic and real datasets. In the synthetic dataset DBN are outperformed by all continuous models. The same happens in the real dataset. The other study, carried out by Gatti et al.[31] addressed the use of CTBN on the diagnosis of acute cardiogenic heart failure while overcoming the main limitations of DBN, referring the interval evidence as a main advantage of CTBN when compared to DBN. There are only presented scenarios, making the model's evaluation subjective.

4.5. Applications of Dynamic Bayesian Belief Networks

The present section will focus on DBBN applied on the medical specialties of Oncology and Others.

In the Oncology field, Watt et al.[52] made use of two datasets, one regarding neuro-oncology and the other osteoarthritis, to apply two networks - Dynamic Bayesian Belief Network (DBBN) and an instant-specific patient model developed using Lazy Bayes Rule (LBR). When the performance on the prediction of a disease was compared, in both datasets the LBR-based DBBN performs better predictive rates over the regular DBBN. On other medical field Watt and Bui[53] constructed a DBBN with the purpose of evaluating both static and dynamic BBN-based models performance in the prediction of knee osteoarthritis diagnosis and compared them to logistic regression models. According to the authors, the results are encouraging, since both static and dynamic belief networks outperformed logistic regression in accuracy for most of variables.

4.6. Applications of other types of Bayesian Networks

The present section of this systematic review will focus on cardiovascular diseases approached by four different Bayesian techniques, such as DBN with Constraint-based fuzzy models, PBN, TNBN, Multilevel BN and MTBN which can be distinguished from this last one because of its capability of analyzing multiple outcomes at multiple time points.

4.6.1. Cardiovascular Diseases

Chiang and Hsu[24] presented a combination between DBN and constraint-based fuzzy models for myocardial infarction classification with 12-lead ECGs. The proposed classifier achieved 86,27% and 78,32% for sensitivity and prediction accuracy respectively. Verduijn et al.[50] on the other hand approached the concept of PBN learned from cardiac surgical patient's data, with hospital mortality as main outcome. The PBN was validated in terms of calibration and discrimination, with the last one being quantified in terms of ROC curve (AUC), and compared to a minimal description length (MDL) network. When comparing the discriminative performance of both networks, PBN had a greater performance than MDL, with higher AUC values. Both networks got poor results in terms of calibration. During the study the authors had some issues on the calibration area which compromised the results. Lappenschaar et al.[33] used a multilevel BN to predict interactions between heart failure and diabetes mellitus, comparing the results obtained with multilevel regression. When the techniques are compared, the multilevel BN obtains AUC values of 0.90 and 0.84 for diabetes mellitus and heart failure respectively; and the multilevel regression (MLR) model obtains AUC values of 0.89 and 0.83 for the same outcomes. With a model similar to this last one, the same major investigator et al.[13] addressed the course of multimorbidity adopting MTBN, investigating the simultaneous progression of six chronic cardiovascular conditions. It was possible to establish associations between comorbidities and risk of cardiovascular disease progression both in quantitative and qualitative analysis, with a novel way of visualization.

4.6.2. Trauma, emergency room and ICU

In the Trauma, emergency room and ICU category, was possible to found two different types of models, a Hierarchical DBN and a comparison between stationary and non-stationary temporal BN. The first one was proposed by Peelen et al.[44] a hierarchical DBN in order to identify temporal patterns, predict future course of disease and test clinical hypotheses regarding patients in an ICU. Also the models should be able to consider dynamics of organ failure at the ICU on two levels: processes regarding an eventual stay at the ICU and the processes related to development or persistence of organ failure in successive days on ICU environment. The model is applied on a case-study of patients admitted to the ICU with severe sepsis. Three Markov models were developed related to clinical outcome hypotheses written in the literature: Model I (severity of organ failure over time), Model II (failure on specific organs), Model III (extension of Model II, distinguishing between persistent organ failure and new organ failure). For all models, the patient can only present three states: patient at the ICU, patient has died, patient has been discharged from the ICU. Since Model I has only one category with seven possible values, it remained as a DBN, while Model II and III became hierarchical DBN. The first model showed an association between organ failure severity and death at ICU in the first 3 days, also finding that after one week the severity-of-illness is no longer significant for the patient's state at the next day. The second model, comparing to the first one, has the capability of being more specific, since it can distinguish between different organs, which one has a biggest impact on ICU stay or discharge. Model III has shown associations between certain organ failures persistence, with some organs contributing for a certain organ failure, but not being associated with its development. All models were evaluated in terms of error rate, AUC and discriminatory performance with all achieving similar results. On the other hand, as stated above, Kayaalp et al.[32] induced a stationary and 33 non-stationary temporal models from the same dataset, comparing the difference in the predictive performance of the models, in order to model efficiency of both types of models and mortality prediction in ICU. AUC for the stationary model was 0.83 and 0.74 for all the non-stationary models. It was possible to see that when the model gets larger it decays its performance. According to the authors, the study proved that non-stationary models can perform as well or even better than stationary models.

4.6.3. Pulmonology

With the attempt of temporal analyze data of patients in ICU who either have VAP or VAP for short, or ICU patients without VAP, Visscher et al.[51] used a Non-repetitive BN, creating three models. However the results from each model were not compared and the gathered information was used to find associations between variables, with no significant results achieved. On another study, Heijden et al.[2] developed predictive models to evaluate chronic obstructive pulmonary disease (COPD) patient's health, in an attempt to facilitate these patients self-management in a home-care setting. After performing an receiver operating characteristic (ROC) analysis all models shown AUC of 0.63. When the TNBN was tested it achieved values of 0.80, 0.75 and 0.24 for AUC, true positive rate (TPR) and false positive rate (FPR) respectively, showing worst performance than the other models, such as temporal augmented naïve bayes (TANB), Aerial and Bootstrap, which, according to the authors, can be explained by limited data that worsens the performance of the network.

4.6.4. Ophthalmology

The Ophthalmology medical specialty, more specifically glaucoma, was addressed by Ceccon et al.[22] using a DSBN to efficiently remove noisy data points and identify and model the key stages in a temporal process, attempting to obtain better performance in classification when compared with both a static and a dynamic BN. The DSBN (AUC – 0.83) outperformed both the BN (AUC – 0.77) and the DBN (AUC – 0.73) performance, also after cross validation, obtaining better accuracy values. With the intent of analyzing classification models to predict whether a certain patient has glaucoma or not. Tucker et al.[48] developed a Spatio-temporal BN classifier, which outperformed the other BN classifiers with a AUC of 0.88. When compared with linear regression and k-nn, the model gets similar to those, but still worst. According to the authors in addition to the good performance, the proposed model has the capability of making explicit all relationships in the network structure and allows to incorporate temporal aspect of the data in a model, combining different types of data.

4.7. Other medical specialties

In an attempt to calculate the probability of Lyme disease on a specific patient, given its clinical history, Olesen et al.[38] developed a BN that incorporates the full clinical history of individual patients. According to the authors when compared to other study the proposed model achieved higher probability of Lyme disease.

5. Discussion

The application of BN in the health-related problems experienced considerable growth and expansion in the last decades. Besides the existence of substantial research regarding the benefits of the implementation and use of the above mentioned techniques on healthcare area, the available evidence is not much specific when it comes to outcomes exploration, conditioning the conclusions retrieved.

Since the beginning of this systematic review it was our aim to arise novel data and content regarding TBN's application on health-related problems, offering information valuable to the clinical area, strengthening the clinical decision support systems. That way, and given the heterogeneity of our results it is of great value the assessment of patterns regarding the benefits and challenges with TBN's implementations.

This study comprised a total of 42 studies from different countries, regarding different studies designs (including descriptive and analytic studies). Carrying different methodological approaches to have an enhanced understanding of the impact and benefits of using these kind of techniques on healthcare data.

As stated above, it was found considerable heterogeneity in the outcomes assessed, as well as in the applied techniques in the studies. Furthermore some studies focused only on a subjective evaluation, turning their assessment less accurate. For this purpose, the results were organized by type of TBN divided by the respective medical specialties of application.

In spite of the variability of the assessed outcomes, our analysis shows common evidence on the use of TBN.

The most commonly found type of BN during this systematic review was DBN, comprising 55% of the studies' total. This particular technique was used with many objectives such as: discover features affecting more prominently the progression of coronary atherosclerosis[27] (like diabetes, cholesterol, family history of coronary disease, among others); identify key role factors towards the progression of atherosclerosis[28]; discover of possible key biomarkers for the diagnosis of hypertension[18]; detection of cardiac arrhythmias[39]; association between patient variables, risk of developing CVD and survival[35]; survival's prediction after ventricular assist device implantation[30]; discovery of possible temporal association of high quality metabolic features for ovarian cancer[17] and colorectal cancer[16] diagnosis; identify factors that dictate oral squamous cell carcinoma progression and to predict potential disease relapses[29]; analyze the impact of prehospital hypotension on systemic acute inflammatory response to blunt trauma[20]; find associations between trauma and inflammatory mediators[19]; model sequential organ failure in adult critically ill patients[4]; capture VAP progression through time[11]; represent clinical variables and the temporal and atemporal relationships underlying insulin and glucose homeostasis[37]; establish relationships between circulating inflammatory mediators and PALF[21]; estimate the risk of fall in elderly[26]; surveillance of patients suffering from renal failure and treated by hemodialysis[46].

For the great majority of these studies it was possible to extract increased value. For example, in the study[21] aiming to establish relationships between circulating inflammatory mediators and PALF, the authors concluded that the network may be used as new biomarkers for predicting outcomes in PALF. Another example is the study[4] that tried to model sequential organ failure in adult critically ill patients, where the according to the authors the study showed the existence of organ failure sequences more likely to occur than others.

Regarding the network's performance it is important to apprise that some studies didn't assessed objective outcomes. In the ones that applied model's evaluation measurements, it is possible to highlight an improvement of 6% in sensitivity on the study approaching detection of cardiac arrhythmias[39]; the study addressing survival's prediction after ventricular assist device implantation reached accuracy values of 96,67% and 93,55% for the two developed DBN[30]; also the study whose aim was to identify factors that dictate oral squamous cell carcinoma progression and to predict potential disease relapses when baseline analysis and first follow-up visit were aggregated towards prediction of relapse probability at second follow-up visit obtained values of 100% in sensitivity, specificity and accuracy, which compared to previous literature shows greater accuracy values in a bigger sample[29].

Concerning DBN with TA, this type of network was modeled for three different objectives, within the medical specialty of CVD, risk assessment[42], diagnosis[41] and prognosis[43] of CAD. Regarding performance, in all studies which the models were evaluated, the DBN with TA outperformed regular DBN for all values, such as sensitivity, specificity, F1score and precision.

Relating to CTBN, it was possible to find model's evaluation in one of the studies with the performance result being favorable for CTBN comparing to DBN, in the study of post-stroke rehabilitation[25].

About DBBN, favorable performance was found in knee osteoarthritis diagnosis with both static and dynamic belief networks outperforming logistic regression in accuracy for most of variables[53].

Addressing other types of BN it's important to highlight a MBN applied to predict interactions between heart failure and diabetes mellitus, comparing the results obtained with MLR, in which MBN slightly outperformed MLR in AUC values, for both heart failure and diabetes[33]. In another study worth noting was applied a MTBN, addressing the course of multimorbidity, considering the simultaneous progression of six chronic cardiovascular conditions, where it was found associations between comorbidities and risk of CVD progression[13]. In the Trauma, emergency room and ICU medical specialty it is essential to point a study that applied a hierarchical DBN in the endeavor of predicting a disease's course in ICU patients, in which the proposed model outperformed DBN, and additionally it showed that some organs contribute for a certain organ failure persistence and, also, it was able to distinguish the impact of different organs on ICU stay or discharge[44].

In summary, this systematic review showed the potential of TBN's use on clinical decision support systems by carrying value on both performance and associations establishments. Addressing performance, it is possible to indicate the superiority of BN containing temporal integration when compared to static BN or other types of techniques in the great majority of studies. It was also possible to stablish important temporal associations between key biomarkers and the progression, risk assessment or diagnosis of some diseases, like hypertension, atherosclerosis, ovarian cancer, systemic inflammation, sepsis, among others.

The resulting evidence unravels information about the large number of techniques derived from BN, which might be used in a future development of clinical decision support systems.

6. Concluding remarks

From the developed work we identified several findings and can now propose some recommendations.

6.1. Main findings

It is possible to highlight the following main findings:

- The found evidence showed the benefit of adding temporal information to Bayesian networks, both in performance and in the discovery of new associations between variables;
- Although the large number of papers published addressing Bayesian networks it was possible to find a lack of application of temporal Bayesian networks on health-related problems;
- In this study it was discovered a sheer number of types of networks developed in the last years.

6.2. Main recommendations

As main recommendations it is possible to propose:

- The use of temporal Bayesian networks applied to health-related problems could stand as a relevant contribute;
- It would be of great relevance the use of more standardized Bayesian network models;
- The use of more homogeneous model's evaluation measures is essential in order to assess the evidence of state of the art methods.

As future work, the application of temporal Bayesian networks in the field of Inflammatory Bowel Disease and Myocardial Infarction is already under development.

7. References

1. Burnside ES: **Bayesian networks: Computer-assisted diagnosis support in radiology**. *Acad Radiol* 2005, **12**:422–430.

2. Van der Heijden M, Velikova M, Lucas PJFF: **Learning Bayesian networks for clinical time series analysis**. *J Biomed Inform* 2014, **48**:94–105.

3. Lucas P: **Bayesian networks in medicine: a model-based approach to medical decision making**. *Proc EUNITE Work Intell …* 2001.

4. Sandri M, Berchialla P, Baldi I, Gregori D, De Blasi RA, Blasi RA De: **Dynamic Bayesian Networks to predict sequences of organ failures in patients admitted to ICU**. *J Biomed Inform* 2014, **48**:106–113.

5. Lucas PJF, Van Der Gaag LC, Abu-Hanna A: **Bayesian networks in biomedicine and health-care**. *Artif Intell Med* 2004, **30**:201–214.

6. Visscher S, Kruisheer EM, Schurink CAM, Lucas PJF, Bonten MJM: **Predicting pathogens causing ventilator-associated pneumonia using a Bayesian network model**. *J Antimicrob Chemother* 2008, **62**:184–188.

7. Young JD, Santos Jr E: **Introduction to temporal Bayesian networks**. In *Online Proceedings of the 1996 Midwest Artificial Intelligence and Cognitive Science Conference*; 1996.

8. Arroyo-Figueroa G, Sucar L: **Temporal network of events for diagnosis and prediction in dynamic domains**. *Appl Intell* 2005, **23**:77–86.

9. Orphanou K, Stassopoulou A, Keravnou E: **Temporal abstraction and temporal Bayesian networks in clinical domains: {A} survey**. *Artif Intell Med* 2014, **60**:133–149.

10. Long W: **Temporal reasoning for diagnosis in a causal probabilistic knowledge base**. *Artif Intell Med* 1996, **8**:193–215.

11. Charitos T, van der Gaag LCCC, Visscher S, Schurink KAMAM, Lucas PJFJFJF: **A dynamic Bayesian network for diagnosing ventilator-associated pneumonia in ICU patients**. *Expert Syst Appl* 2009, **36**:1249–1258.

12. Augusto JC: **Temporal reasoning for decision support in medicine**. *Artif Intell Med* 2005, **33**:1–24.

13. Lappenschaar M, Hommersom A, Lucas PJF, Lagro J, Visscher S, Korevaar JC, Schellevis FG: **Multilevel temporal Bayesian networks can model longitudinal change in multimorbidity**. *J Clin Epidemiol* 2013, **66**:1405–1416.

14. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P,

Stewart LA: **Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement**. *Syst Rev* 2015, **4**:1.

15. Whiting PF, Rutjes AWS, Westwood ME, Mallet S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne J a C, Bossuyt PMM: **Research and Reporting Methods Accuracy Studies**. *Ann Intern Med* 2011, **155**:529–536.

16. Akutekwe A, Seker H: **Particle swarm optimization-based bio-network discovery method for the diagnosis of colorectal cancer**. In *Proceedings - 2014 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2014*. School of Computer Science and Informatics, Faculty of Technology, De Montfort University, Leicester, United Kingdom; 2014:8–13.

17. Akutekwe A, Seker H: **Two-stage computational bio-network discovery approach for metabolites: Ovarian cancer as a case study**. In *2014 IEEE-EMBS International Conference on Biomedical and Health Informatics, BHI 2014*. Bio-Health Informatics Research Group, Faculty of Technology, De Montfort University, Leicester, LE1 9BH, United Kingdom; 2014:97–100.

18. Akutekwe A, Seker H: **A hybrid dynamic Bayesian network approach for modelling temporal associations of gene expressions for hypertension diagnosis**. *2014 36th Annu Int Conf IEEE Eng Med Biol Soc EMBC 2014* 2014, **2014**:804–807.

19. Almahmoud K, Namas RA, Abdul-Malak O, Zaaqoq AM, Zamora R, Zuckerbraun BS, Sperry J, Peitzman AB, Billiar TR, Vodovotz Y, K. A, R.A. N, O. A-M, A.M. Z, R. Z, B.S. Z, J. S, A.B. P, T.R. B, Y. V: **Impact of injury severity on dynamic inflammation networks following blunt trauma**. *Shock* 2015, **44**:101–109.

20. Almahmoud K, Namas RA, Zaaqoq AM, Abdul-Malak O, Namas R, Zamora R, Sperry J, Billiar TR, Vodovotz Y: **Prehospital hypotension is associated with altered inflammation dynamics and worse outcomes following blunt trauma in humans**. *Crit Care Med* 2015, **43**:1395–1404.

21. Azhar N, Ziraldo C, Barclay D, Rudnick DA, Squires RH, Vodovotz Y: **Analysis of serum inflammatory mediators identifies unique dynamic networks associated with death and spontaneous survival in pediatric acute liver failure**. *PLoS One* 2013, **8**:e78202.

22. Ceccon S, Garway-Heath D, Crabb DP, Tucker A: **The Dynamic Stage Bayesian Network: Identifying and Modelling Key Stages in a Temporal Process**. In *IDA. Volume 7014*. Springer; 2011:101–112. [*Lecture Notes in Computer Science*]

23. Chang R, Stetter M, Brauer W: **Quantitative Inference by Qualitative Semantic Knowledge Mining with Bayesian Model Averaging**. *IEEE Trans Knowl Data Eng* 2008, **20**:1587–1600.

24. Chiang Y-Y, Hsu W-H: **Integrating Dynamic Bayesian Networks and Constraint-based Fuzzy Models for Myocardial Infarction Classification with 12-lead Ecgs**. *2010 Conf Precis Electromagn Meas Cpem* 2010:IEEE.

25. Codecasa D, Stella F: **Learning continuous time Bayesian network classifiers**. *Int J Approx Reason* 2014, **55**:1728–1746.

26. Cuaya G, Munoz-Melendez A, Nunez Carrera L, Morales EF, Quinones I, Perez AI, Alessi A: **A dynamic Bayesian network for estimating the risk of falls from real gait data.** *Med Biol Eng Comput* 2013, **51**:29–37.

27. Exarchos KP, Exarchos TP, Bourantas C V, Papafaklis MI, Naka KK, Michalis LK, Parodi O, Fotiadis DI: **Prediction of Coronary Atherosclerosis Progression using Dynamic Bayesian Networks**. *2013 35th Annu Int Conf Ieee Eng Med Biol Soc* 2013, **2013**:3889–3892.

28. Exarchos KP, Carpegianni C, Rigas G, Exarchos TP, Vozzi F, Sakellarios A, Marraccini P, Naka K, Michalis L, Parodi O, Fotiadis DI: **A multiscale approach for modeling atherosclerosis progression**. *IEEE J Biomed Heal Informatics* 2015, **19**:709–719.

29. Exarchos KP, Goletsis Y, Fotiadis DI: **A multiscale and multiparametric approach for modeling the progression of oral cancer.** *BMC Med Inform Decis Mak* 2012, **12**:136.

30. Exarchos TP, Rigas G, Goletsis Y, Stefanou K, Jacobs S, Trivella M-G, Fotiadis DI: **A dynamic Bayesian network approach for time-specific survival probability prediction in patients after ventricular assist device implantation**. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*. *Volume 2014*. Dept. of Biomedical Research, Institute of Molecular Biology and Biotechnology, FORTH University of Ioannina, Greece, United States; 2014:3172–3175.

31. Gatti E, Luciani D, Stella F: **A continuous time Bayesian network model for cardiogenic heart failure**. *Flex Serv Manuf J* 2012, **24**:496–515.

32. Kayaalp M, Cooper GF, Clermont G: **Predicting ICU mortality: a comparison of stationary and nonstationary temporal models.** *Proc AMIA Symp* 2000:418–422.

33. Lappenschaar M, Hommersom A, Lucas PJF, Lagro J, Visscher S: **Multilevel Bayesian networks for the analysis of hierarchical health care data**. *Artif Intell Med* 2013, **57**:171–183.

34. Marini S, Trifoglio E, Barbarini N, Sambo F, Di Camillo B, Malovini A, Manfrini M, Cobelli C, Bellazzi R: **A Dynamic Bayesian Network model for long-term simulation of clinical complications in type 1 diabetes.** *J Biomed Inform* 2015.

35. Marshall AH, Hill LA, Kee F: **Continuous Dynamic Bayesian networks for predicting survival of ischaemic heart disease patients**. In *Proceedings - IEEE Symposium on Computer-Based Medical Systems*. Centre for Statistical Science and Operational Research (CenSSOR), Queen's University Belfast, Belfast, BT7 1NN, United Kingdom: {IEEE} Computer Society; 2010:178–183.

36. Nachimuthu SK, Haug PJ: **Early detection of sepsis in the emergency department using Dynamic Bayesian Networks.** *AMIA Annu Symp Proc* 2012, **2012**:653–662.

37. Nachimuthu SK, Wong A, Haug PJ: **Modeling Glucose Homeostasis and Insulin Dosing in an Intensive Care Unit using Dynamic Bayesian Networks.** *AMIA Annu Symp Proc* 2010, **2010**:532–536.

38. Olesen KG, Hejlesen OK, Dessau R, Beltoft I, Trangeled M: **Diagnosing Lyme disease - Tailoring patient specific Bayesian networks for temporal reasoning**. In *Probabilistic Graphical Models*; 2006:231–238.

39. Oliveira LSC De, Andreão R V, Sarcinelli-filho M: **Premature Ventricular Beat Classification Using a Dynamic Bayesian Network**. *33rd Annu Int Conf IEEE EMBS* 2011, **2011**:4984–4987.

40. Orphanou K, Keravnou E, Moutiris JA: **Integration of Temporal Abstraction and Dynamic Bayesian Networks in Clinical Systems. {A} preliminary approach**. In *ICCSW. Volume 28*. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany; 2012:102–108. [*OASICS*]

41. Orphanou K, Stassopoulou A, Keravnou E: **Integration of Temporal abstraction and Dynamic Bayesian Networks for Coronary Heart Diagnosis**. In *{STAIRS} 2014 - Proceedings of the 7th European Starting {AI} Researcher Symposium, Prague, Czech Republic, August 18-22, 2014. Volume 264*. Edited by Endriss U, Leite J. {IOS} Press; 2014:201–210. [*Frontiers in Artificial Intelligence and Applications*]

42. Orphanou K, Stassopoulou A, Keravnou E: **Risk Assessment for Primary Coronary Heart Disease Event Using Dynamic Bayesian Networks**. In *AIME. Volume 9105*. Springer; 2015:161–165. [*Lecture Notes in Computer Science*]

43. Orphanou K, Stassopoulou A, Keravnou E: **DBN-extended: A Dynamic Bayesian network model extended with temporal abstractions for coronary heart disease prognosis.** *IEEE J Biomed Heal informatics* 2015.

44. Peelen L, de Keizer NFNF, de Jonge E, Bosman R-J, Abu-Hanna A, Peek N: **Using hierarchical dynamic Bayesian networks to investigate dynamics of organ failure in patients in the Intensive Care Unit**. *J Biomed Inform* 2010, **43**:273–286.

45. Peeling E, Tucker A: **Making time: Pseudo time-series for the temporal analysis of cross section data**. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 2007, **4723 LNCS**:184–194.

46. Rose C, Smaili C, Charpillet F: **A dynamic Bayesian network for handling uncertainty in a decision support system adapted to the monitoring of patients treated by hemodialysis**. *Ictai 2005 17th Ieee Int Conf Tools with Artif Intell Proc* 2005:IEEE Comp Soc; Informat Technol Res Inst; Wright S.

47. Sebastiani P, Mandl KD, Szolovits P, Kohane IS, Ramoni MF: **A Bayesian dynamic model for influenza surveillance.** *Stat Med* 2006, **25**:1803–1816.

48. Tucker A, Vinciotti V, Liu X, Garway-Heath D: **A spatio-temporal Bayesian network classifier for understanding visual field deterioration**. *Artif Intell Med* 2005, **34**:163–177.

49. van Gerven MAJ, Taal BG, Lucas PJF: **Dynamic Bayesian networks as prognostic models for clinical patient management**. *J Biomed Inform* 2008, **41**:515–529.

50. Verduijn M, Rosseel PMJ, Peek N, de Jonge E, de Mol BAJM: **Prognostic Bayesian networks. II: An application in the domain of cardiac surgery**. *J Biomed Inform* 2007, **40**:619–630.

51. Visscher S, Lucas P, Flesch I, Schurink K: **Using temporal context-specific independence information in the exploratory analysis of disease processes**. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 2007, **4594**:87–96.

52. Watt E, Sayre JW, Bui AAT: **Applying an instance-specific model to longitudinal clinical data for prediction**. *Proc - 2011 1st IEEE Int Conf Healthc Informatics, Imaging Syst Biol HISB 2011* 2011:81–88.

53. Watt E, Bui AAT: **Evaluation of a Dynamic Bayesian Belief Network to Predict Osteoarthritic Knee Pain Using Data from the Osteoarthritis Initiative**. In *AMIA*. University of California Los Angeles, Los Angeles, CA, USA.: AMIA; 2008:788–792.

8. Annexes

Annex 1 - PROSPERO's Committee answer

CRD-REGISTER <irss505@york.ac.uk>
para mim

inglês > português Traduzir mensagem

Dear Mr Martins

Thank you for submitting details of your systematic review *Evidence of Temporal Bayesian Networks applications for health-related problems* to the PROSPERO register.

We regret that the information you have provided fails to meet the requirements for inclusion. Specifically, this is a methodology review which has no directly health related outcomes.

We hope that this will not discourage you from registering your next systematic review at the protocol stage, with PROSPERO.

Yours sincerely

James Christie

PROSPERO Administrator
Centre for Reviews and Dissemination
University of York
York YO10 5DD
t: +44 (0) 1904 321040
f: +44 (0) 1904 321041
e: CRD-register@york.ac.uk
www.york.ac.uk/inst/crd